

## **EVALUACIÓN SUBJETIVA DE LA DIFERENCIACIÓN ENTRE LAS RESPUESTAS IMPULSIVAS MEDIDAS Y SIMULADAS EN DIFERENTES RECINTOS**

PACS: 43.55.Ka

Ortiz Mora, José Ignacio; Pedrero González, Antonio; Díaz Sanchidrián, César  
Grupo de investigación en Acústica Arquitectónica. Universidad Politécnica de Madrid.  
E.T.S. de Arquitectura. Avenida Juan de Herrera nº4 28040 Madrid, España.  
Tel:+34 913364249 Fax:+34 913366554  
[antonio.pedrero@upm.es](mailto:antonio.pedrero@upm.es)

### **ABSTRACT**

This work aims the subjective assessment of the differentiation between auralization signals with impulse responses simulated and impulse responses measured at different rooms. To this end, we have developed a software tool that generates the corresponding auralizations, provides a user interface to perform a subjective listening test and stores the results of the test.

In this project we present the results of a listening test performed for 58 listeners, using different test signals from the impulse responses measured and simulated in six rooms with different acoustic characteristics.

### **RESUMEN**

Este trabajo tiene como objeto la evaluación subjetiva de la diferenciación entre señales auralizadas con respuestas impulsivas simuladas y con respuestas impulsivas medidas en diferentes recintos. Para ello, se ha desarrollado una herramienta software que, genera las auralizaciones correspondientes, proporciona una interfaz de usuario para la realización de un test subjetivo de escucha y almacena los resultados de dicho test.

Se presentan los resultados de un test de escucha realizado a 58 oyentes, utilizando diferentes señales de prueba, a partir de las respuestas impulsivas simuladas y medidas en seis recintos con características acústicas distintas.

## **INTRODUCCIÓN**

Las técnicas de Realidad Acústica Virtual son unas herramientas cada vez más usadas en acústica de salas. La elaboración de modelos acústicos digitales de los espacios, reales o imaginarios, permite la auralización del sonido en dichos espacios. Estas auralizaciones se pueden usar para valorar las características acústicas de un recinto, para identificar posibles problemas acústicos o como herramienta de demostración del comportamiento acústico de los espacios en su fase de diseño. Por esto, es necesario asegurarse de que las auralizaciones tienen suficiente calidad para representar fielmente el sonido real de los espacios modelados.

Hasta la fecha, no existe un procedimiento establecido para valorar la calidad de las auralizaciones, aunque el método más utilizado hasta la fecha ha consistido en la realización de pruebas subjetivas en las que se comparan los sonidos reales grabados en los recintos con los obtenidos mediante auralización [1][2].

Por otra parte, el criterio usualmente utilizado para la validación de los modelos acústicos digitales de los recintos es la comparación de los valores de las magnitudes acústicas obtenidas en las mediciones acústicas realizadas in situ con los valores de las mismas magnitudes acústicas que se obtienen en el modelo digital. Para cuantificar las diferencias se utilizan los valores de Umbral Diferencial, JND (Just Noticeable Differences) expresados en la norma ISO 3382-1[3], que se definen como la mínima diferencia en el valor de cada una de las magnitudes acústicas que es detectable perceptualmente. Aunque en rigor los valores de JND se refieren a la banda de frecuencia de 1 KHz, en ocasiones, los JND se aplican para el promedio de frecuencias medias (500 y 1000 Hz), o incluso para el promedio de todas las bandas de frecuencia evaluadas. En cuanto a las magnitudes acústicas utilizadas en la validación de los modelos digitales, en algunos casos se utiliza exclusivamente el parámetro  $T_{30}$ , en otros se utiliza una combinación de los parámetros expresados en la norma ISO 3382-1.

El objetivo de este trabajo es evaluar si los oyentes diferencian auralizaciones realizadas con respuestas impulsivas reales de recintos y las realizadas con respuestas impulsivas obtenidas de modelos acústicos digitales de los mismos recintos, debidamente validados. También interesa conocer hasta qué punto la diferencia entre los valores de las magnitudes acústicas, en términos de JND, obtenidas a partir de respuestas impulsivas medidas y simuladas, se corresponde con la diferencia subjetiva de las auralizaciones efectuadas con las mismas respuestas impulsivas. Para ello se ha realizado un test subjetivo de escucha en el que se han utilizado respuestas impulsivas medidas y simuladas en diversos recintos con diferentes características acústicas. Para cada respuesta impulsiva se han realizado auralizaciones con tres diferentes señales sonoras.

## **DISEÑO DEL TEST DE ESCUCHA**

Para evaluar la similitud entre los espacios reales y los modelados a partir de programas de simulación acústica, se ha efectuado una evaluación subjetiva mediante un test de escucha que analiza la diferenciación entre las señales auralizadas que se han generado con respuestas impulsivas obtenidas en las mediciones realizadas en un determinado recinto con respecto a las que se han generado con respuestas impulsivas obtenidas mediante la simulación digital del mismo recinto.

A falta de un procedimiento específico para la realización de test subjetivos de escucha en el campo de la acústica de salas, se han seguido los criterios de la Recomendación ITU-R-BS 1116-1 (1997)[4], que se utiliza para evaluar pequeñas degradaciones en señales de audio. El método

utilizado es el denominado «triple estímulo doblemente ciego con referencia oculta». En este método, se presentan al participante tres estímulos sonoros («1», «2» y «3»), dos de los cuales son iguales y el tercero es ligeramente diferente. El participante debe identificar cuáles son las señales iguales y cuál la diferente. La asignación numérica se realiza aleatoriamente para cada triada de señales.

Para facilitar la ejecución del test y automatizar el registro de los resultados se ha creado una herramienta software basada en MATLAB. La herramienta se ha basado en el software WhisPER [5], desarrollado por el Audio Communication Group de la Universidad de Berlín, del que se ha modificado la interfaz gráfica y se han incluido las funciones correspondientes para cumplir con los objetivos y requisitos de este trabajo.

Como se ha comentado anteriormente el test de escucha consiste en evaluar la diferencia entre dos pistas auralizadas, utilizando un triple estímulo doblemente ciego con referencia oculta. El oyente tendrá que identificar cuáles de los tres estímulos sonoros ha percibido como iguales y cuál ha sido el grado de diferencia encontrada con el tercer estímulo, en el caso de que la haya. La diferencia se puede valorar como “Pequeña”, “Moderada” o “Grande”. Si el participante no encuentra diferencias entre las tres señales que componen una triada, el software inhabilita el módulo de valoración de la diferencia y finaliza la evaluación de dicha triada de señales. Cuando el participante aprecia señales diferenciadas, se inicia una segunda parte del test en la que se le pregunta por las diferencias percibidas en función de tres parámetros acústicos: reverberación, proximidad e inteligibilidad/claridad. La Figura 1 muestra la interfaz de usuario diseñada para el test subjetivo.

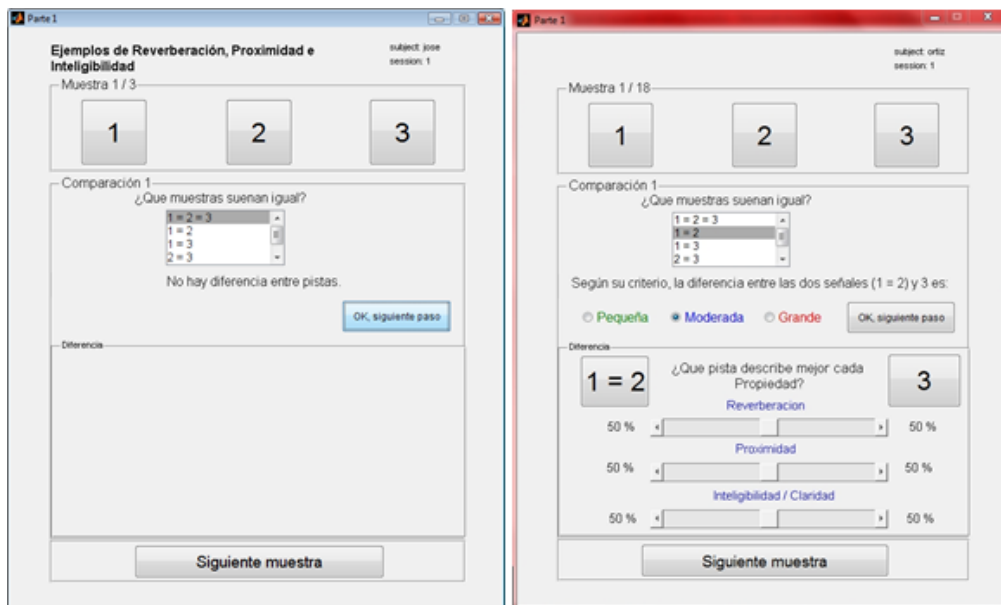


Figura 1. Pantalla ejemplo del test de escucha cuando las tres pistas se perciben iguales y cuando se perciben dos iguales y una diferente

En esta evaluación subjetiva han evaluado 18 parejas de respuestas impulsivas (medidas y simuladas), correspondientes a 6 recintos diferentes con tres puntos distintos en cada recinto. Los

recintos utilizados tienen diferentes características acústicas, como se puede observar en la Tabla 1.

Magnitud	Recinto 1	Recinto 2	Recinto 3	Recinto 4	Recinto 5	Recinto 6
EDT [s]	2,61	5,59	1,46	0,98	2,19	1,39
$T_{30}$ [s]	2,79	6,44	1,30	0,83	2,17	1,21
$T_s$ [ms]	206	380	124	75	181	118
$C_{80}$ [dB]	-4,3	-4,5	-1,5	2,7	-3,4	-1,0
$D_{50}$ [-]	0,21	0,31	0,26	0,49	0,18	0,25
RASTI [-]	0,39	0,38	0,49	0,61	0,43	0,50

Tabla 1. Valores de las magnitudes acústicas para frecuencias medias, obtenidos mediante mediciones in situ en los recintos utilizados

La Figura 2 muestra los valores del tiempo de reverberación  $T_{30}$ , en bandas de octava, obtenidos en las mediciones y en las simulaciones de los distintos recintos.

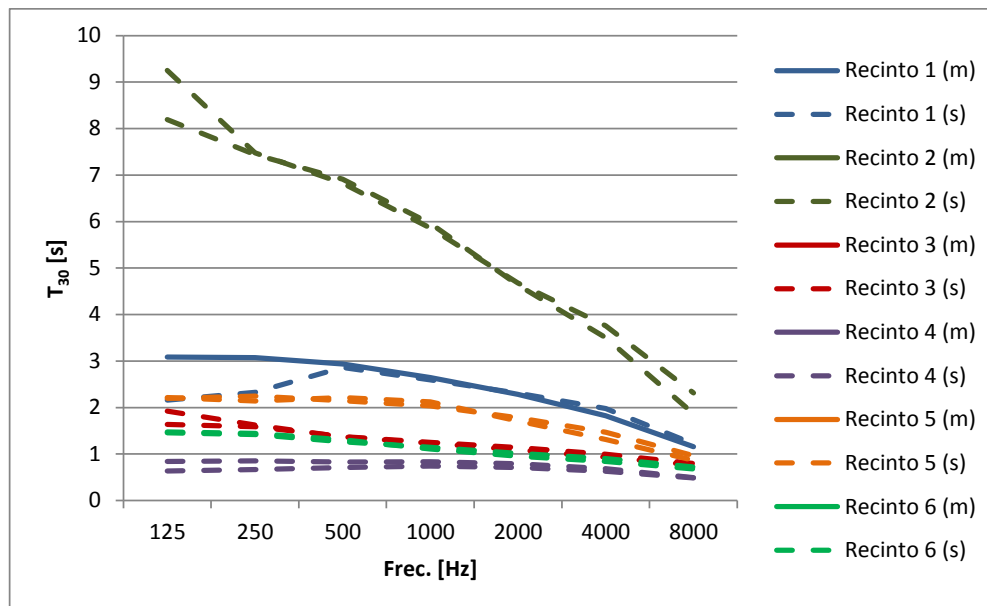


Figura 2. Tiempos de reverberación,  $T_{30}$ , medidos (m) y simulados (s) en los diferentes recintos

Cada una de las respuestas impulsivas se ha convolucionado con tres señales de audio diferentes: una señal con voz hablada (Audio 1), una señal con sonido de palmadas (Audio 2) y una señal musical (Audio 3). En total, se han generado 54 parejas de auralizaciones (18 parejas de respuestas impulsivas por tres señales de audio). Con cada pareja de auralizaciones se confecciona una triada, duplicando alguna de las dos señales resultantes de la auralización. Por lo tanto, se dispone de 54 triadas para experimentar con el test de escucha.

Dado que se recomienda que la duración del test no exceda los 20 minutos, se ha limitado el número de señales a evaluar por cada uno de los participantes a 18 triadas. Por esta razón, las 54 triadas disponibles se han dividido en tres sesiones con 18 triadas cada una. Cada sesión incluye

auralizaciones de los tres puntos de cada recinto, con una distribución similar en cuanto a las señales de audio utilizadas en la auralización.

Para su tratamiento matemático, las respuestas de los participantes se han codificado asignando un número a la opción elegida. En primer lugar se ha evaluado si la discriminación entre señales iguales y señal diferente es correcta. En tal caso, el código asignado ha sido “1”, y en caso de que el participante no haya determinado correctamente cuál es la señal diferente y cuáles son las correctas el código asignado ha sido “0”. Posteriormente se ha codificado el grado de similitud apreciado entre las señales diferentes, asignando un “0” cuando el participante no distingue entre ninguna de las tres señales, “1” cuando la diferencia entre las señales iguales con respecto a la diferente es calificada como “Pequeña”, “2” si se califica como “Moderada” y “3” si el oyente considera que la diferencia entre señales es “Grande”.

Para desechar resultados anómalos, se ha realizado un seguimiento de los participantes durante la realización del test. Además, a los resultados obtenidos se les aplicó un test de Grubs para descartar valores aberrantes.

## RESULTADOS OBTENIDOS

En este trabajo se analizan los resultados de la primera parte del test de escucha llevado a cabo.

En el experimento han participado 58 oyentes de diferentes edades y de ambos sexos. Cada de ellos ha realizado una sesión del test de escucha. En la siguiente tabla se muestran los promedios de edades de los participantes en cada una de las tres sesiones y en total, además del número de hombres y mujeres participantes en cada sesión.

Sesión	Edad Promedio	Nº Hombres	Nº Mujeres
1	34,2	11	9
2	31,6	9	10
3	33,0	9	10
Total	32,9	29	29

Tabla 2. Edades promedio y distribución por sexos de las sesiones del test de escucha.

La Tabla 3 presenta el número de fallos encontrados en la discriminación entre señales iguales y señales diferentes. Cabe destacar que no se ha dado ningún caso en el que señales iguales, se hayan considerado diferentes. Lo que sí ha sucedido es que auralizaciones distintas se han considerado como iguales.

Recinto	Pareja de respuestas impulsivas	Fallos totales	Fallos por Sesión			Fallos por Audio		
			Sesión 1	Sesión 2	Sesión 3	Audio 1	Audio 2	Audio 3
1	1	1	0	0	1	0	0	1
	2	4	0	3	1	1	0	3
	3	4	4	0	0	0	0	4
2	4	2	1	1	0	1	1	0
	5	1	0	0	1	1	0	0
	6	0	0	0	0	0	0	0
3	7	0	0	0	0	0	0	0
	8	1	0	1	0	0	0	1

	9	2	1	1	0	1	0	1
4	10	2	1	0	1	1	0	1
	11	1	0	0	1	1	0	0
	12	2	1	1	0	1	0	1
	13	1	0	0	1	0	0	1
5	14	1	0	1	0	0	0	1
	15	0	0	0	0	0	0	0
	16	0	0	0	0	0	0	0
6	17	1	0	0	1	1	0	0
	18	0	0	0	0	0	0	0
Suma		23	8	8	7	8	1	14
% Fallos		2,2	2,2	2,3	2,0	2,3	0,3	4,0

Tabla 3. Número de fallos en la discriminación de señales

Los resultados en cuanto a la catalogación de las diferencias encontradas entre las señales consideradas distintas por los participantes se muestran en la Figura 3 y en la Tabla 4. La figura 3 presenta el porcentaje de asignación de cada uno de intervalos de calificación para cada triada. Por su parte, la Tabla 4 indica, para cada triada, el valor medio del valor de diferencia asignado ( $\bar{X}$ ), la desviación típica ( $\sigma$ ) y la moda (Mo).

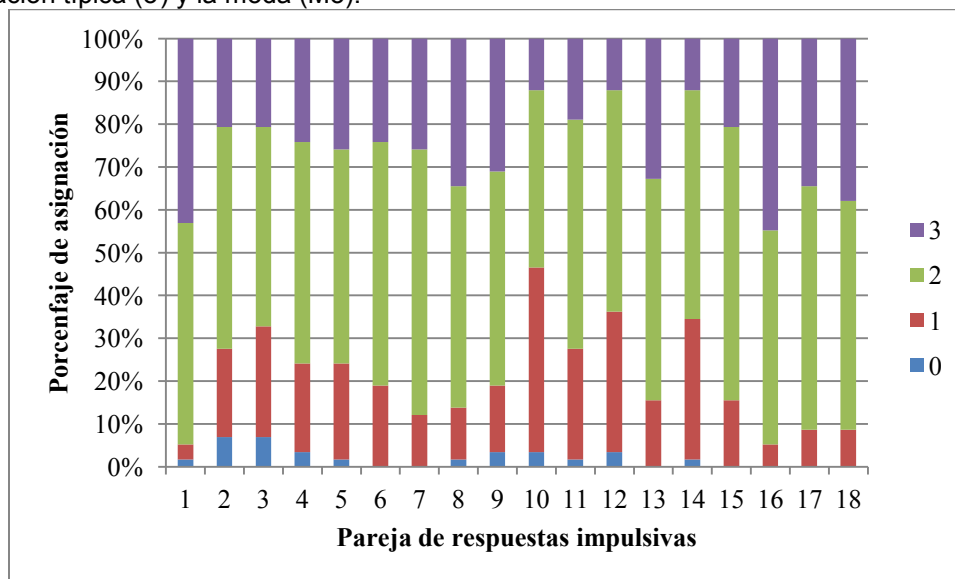


Figura 3. Porcentaje de asignación de cada uno de los intervalos de calificación de las diferencias: 0 indica "Sin diferencias", 1 "Pequeña", 2 "Moderada" y 3 "Grande".

Pareja de respuestas impulsivas	Total			Audio 1			Audio 2			Audio 3		
	$\bar{X}$	$\sigma$	Mo	$\bar{X}$	$\sigma$	Mo	$\bar{X}$	$\sigma$	Mo	$\bar{X}$	$\sigma$	Mo
1	2,4	0,64	2	2,4	0,49	2	2,7	0,48	3	2,1	0,78	2
2	1,9	0,83	2	1,8	0,79	2	2,4	0,60	2	1,4	0,76	2

3	1,8	0,85	2	2,4	0,50	2	2,2	0,54	2	0,9	0,55	1
4	2,0	0,77	2	1,8	0,83	2	1,6	0,68	2	2,5	0,51	2
5	2,0	0,75	2	1,8	0,85	2	2,4	0,60	2	1,8	0,63	2
6	2,1	0,66	2	1,9	0,71	2	1,9	0,62	2	2,3	0,64	2
7	2,1	0,61	2	1,9	0,49	2	2,6	0,51	3	2,0	0,58	2
8	2,2	0,71	2	1,9	0,52	2	2,7	0,47	3	1,9	0,81	2
9	2,1	0,78	2	1,9	0,81	2	2,4	0,51	2	2,0	0,89	2
10	1,6	0,75	1	1,6	0,68	2	1,8	0,85	1	1,5	0,70	1
11	1,9	0,72	2	1,7	0,73	2	2,3	0,72	2	1,7	0,58	2
12	1,7	0,72	2	1,8	0,79	2	1,8	0,69	2	1,6	0,69	2
13	2,2	0,68	2	2,2	0,70	2	2,4	0,61	2	1,9	0,66	2
14	1,8	0,68	2	1,9	0,62	2	1,7	0,66	2	1,6	0,76	2
15	2,1	0,60	2	2,2	0,54	2	2,0	0,67	2	2,0	0,60	2
16	2,4	0,59	2	2,5	0,69	3	2,4	0,60	2	2,3	0,48	2
17	2,3	0,61	2	2,2	0,50	2	2,2	0,75	2	2,5	0,51	2
18	2,3	0,62	2	2,3	0,58	2	2,3	0,56	2	2,3	0,73	3

Tabla 4. Valor medio, desviación típica y moda de la calificación de las diferencias.

## ANÁLISIS DE LOS RESULTADOS

Con respecto a la discriminación entre las señales iguales y la señal diferencia, los resultados muestran un bajo porcentaje de fallos, en torno al 2 %, que se mantiene aproximadamente constante para las tres sesiones de las que consta el test. Sin embargo, existen diferencias apreciables en cuanto al porcentaje de fallos en discriminación en función del tipo de señal de audio utilizada en la auralización. La mejor discriminación se consigue con la señal Audio 2, que consiste en un sonido de palmadas, con unas componentes impulsivas destacadas. En este caso el porcentaje de fallos es 0,3, casi diez veces inferior al valor medio de la prueba. La señal de voz hablada (Audio 1) tiene un porcentaje de fallos similar al valor medio, mientras que la peor discriminación se obtiene con señal musical (Audio 3), con la que se duplica el porcentaje de fallos promedio.

En cuanto a la calificación de las diferencias detectadas, se observa una gran dispersión entre la valoración de los distintos participantes. De las 18 parejas de auralizaciones, 11 reciben las cuatro calificaciones posibles, con diferentes porcentajes, y las siete parejas restantes son calificadas con tres de los posibles intervalos de calificación. La calificación más frecuente de la diferencia es "Moderada", que se codifica con el número 2.

También se intenta comprobar si existe algún tipo de correlación entre las diferencias de las magnitudes acústicas de las respuestas impulsivas medidas y simuladas y la calificación de la diferencia obtenida en el test de escucha. Con este objetivo, para cada punto analizado, se ha comparado la respuesta impulsiva medida y simulada, calculándose la diferencia en los valores de los diferentes parámetros acústicos en términos de JND, para cada una de las bandas de octava con frecuencias centrales entre 125 y 8000 Hz. Para obtener un valor único de JND por cada punto evaluado se han realizado diferentes promedios de los valores obtenidos, en función de las magnitudes acústicas y las bandas de frecuencia consideradas en cada caso. Denominamos "Promedio" a la media aritmética de los valores de diferencias en términos de JND obtenidos para EDT,  $T_{30}$ ,  $T_S$ ,  $C_{80}$ ,  $D_{50}$ , en todas las bandas de frecuencia, y RASTI. "1KHz" es el valor promedio de las diferencias obtenidas para EDT,  $T_{30}$ ,  $T_S$ ,  $C_{80}$  y  $D_{50}$ , para la banda de octava de 1 KHz. "MID" representa el valor de las diferencias de los parámetros anteriores promediados en las bandas de



500 y 1000 Hz. “T30 1KHz” t “T30 MID” son las diferencias de dicho parámetro en la banda de 1KHz y en el promedio de 500 y 1000 Hz, respectivamente. En la Tabla 2 se muestran las diferencias, en términos de JND, para cada punto de cada recinto.

		Diferencias [JND]				
		Promedio	1 KHz	MID	T <sub>30</sub> 1KHz	T <sub>30</sub> MID
Recinto 1	Punto 1	10,08	11,34	10,79	1,47	1,57
	Punto 2	6,66	4,53	5,56	0,12	0,24
	Punto 3	6,13	4,56	5,54	0,46	0,47
Recinto 2	Punto 1	8,62	4,28	3,94	0,59	0,28
	Punto 2	3,65	2,14	3,11	0,35	0,43
	Punto 3	3,64	1,61	1,55	0,12	0,18
Recinto 3	Punto 1	5,39	5,99	5,34	0,75	0,81
	Punto 2	6,65	6,26	5,80	0,15	0,83
	Punto 3	3,08	2,21	2,13	0,18	0,61
Recinto 4	Punto 1	4,61	1,91	3,21	3,26	4,66
	Punto 2	3,17	3,52	2,57	1,91	2,08
	Punto 3	2,10	1,67	1,26	1,86	1,17
Recinto 5	Punto 1	5,43	5,30	4,80	0,32	0,52
	Punto 2	1,65	0,62	0,39	1,13	0,52
	Punto 3	2,32	1,05	0,42	1,14	1,20
Recinto 6	Punto 1	3,06	4,03	3,33	0,83	0,14
	Punto 2	2,44	1,70	1,46	1,03	0,34
	Punto 3	1,92	1,23	1,24	0,39	0,25

Tabla 5. Diferencias entre los valores de las magnitudes acústicas obtenidas a partir de las respuestas impulsivas medidas y simuladas.

Se han calculado las correlaciones entre las diferencias en JND y los valores de calificación obtenidos en el test subjetivo. En general, los coeficientes de determinación encontrados son muy inferiores a la unidad, lo que indica que no existe una relación lineal entre las diferencias basadas en los índices objetivos y las basadas en el test subjetivo. Las mejores correlaciones se dan para las auralizaciones realizadas con la señal Audio 2. De todos casos, el coeficiente de determinación más alto ( $R^2=0,42$ ) se obtiene con esta señal de audio y para el promedio de los JND de todas las magnitudes acústicas en la banda de octava de 1KHz.

## CONCLUSIONES

Se ha diseñado una herramienta automática para realizar test subjetivos de escucha basados en el método de triple estímulo doblemente ciego con referencia oculta. Dicha herramienta se ha utilizado en una prueba en la que se persigue evaluar la discriminación entre auralizaciones realizadas con señales impulsivas medidas y respuestas impulsivas simuladas en distintos recintos.

Los resultados muestran que la mayoría de los participantes encuentran diferencias entre ambos tipos de auralizaciones. La diferenciación depende de una manera importante de la señal de audio que se utiliza como base de la auralización, siendo inferior cuando se utilizan señales musicales



que para señales de voz hablada o impulsivas. Se observa que no se ha dado ningún caso en el que señales iguales, se hayan considerado diferentes. Los fallos de diferenciación han consistido en que auralizaciones distintas se han considerado iguales.

En un primer análisis de los resultados, no se ha encontrado una correlación clara entre el grado de diferenciación resultante del test de escucha y las diferencias entre los valores de las magnitudes acústicas derivadas de las respuestas impulsivas, expresadas en términos de JND.

## **AGRADECIMIENTOS**

Los autores agradecen al Audio Communication Group de la Universidad de Berlín y, en particular, a Alexander Lindau por permitirnos el uso del código de WhisPER para desarrollar el test de escucha. Nuestro agradecimiento también a todas las personas que participaron en la evaluación subjetiva y ayudaron en la realización del trabajo.

## **REFERENCIAS**

- [1] T. Lokki, H. Järveläinen, "Subjective evaluation of auralization of physics-based room acoustics modelling", Proceedings of the 2001 International Conference on Auditory Display, Espoo, Finland, July 29-August 1, 2001
- [2] J.H. Rindel, C.L. Christensen. "Room acoustic simulation and auralization - how close can we get to the real room?", Proceedings of Eighth Western Pacific Acoustics Conference, Melbourne, Australia, April 7-9 2003.
- [3] ISO 3382-1:2009. Acoustics -- Measurement of room acoustic parameters -- Part 1: Performance spaces. International Organization for Standardization.
- [4] Recomendación ITU-R-BS1116-1. Métodos para la evaluación subjetiva de pequeñas degradaciones en los sistemas de audio incluyendo los sistemas de sonido multicanal.
- [5] S. Ciba, A. Wlodarski, H.J. Maempel, "WhisPER- A new tool performing listening test", AES 126th Convention, Munich, Germany, May 7–10, 2009